Probabilidade e Estatística Matemática

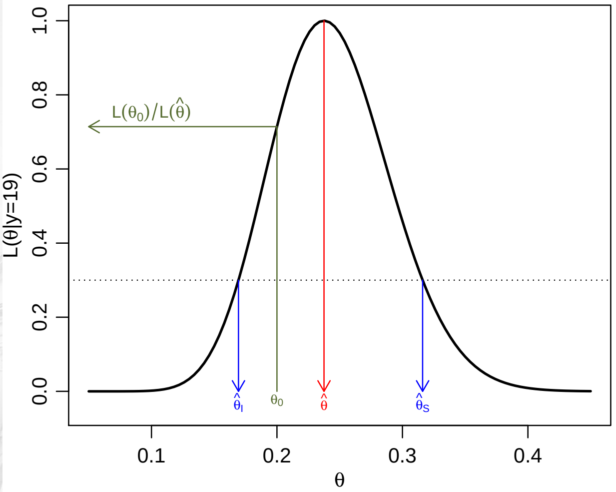
Ano 2022 - Módulo 4 - Aula 1 (Parte 2/2)

Paulo Justiniano Ribeiro Jr

Programa de Pós Graduação
Métodos Numéricos em Engenharia
Universidade Federal do Paraná

10 de agosto de 2022

Função de Verossimilhança



- máximo: estimativa pontual,
- pontos de corte: estimativa intervalar,
- valor em um ponto: razoabilidade do valor.

DEST Departamento de Estatística

Expressão da função de verossimilhança (I)

V.A. observável discreta (não há ambiguidade)

$$L(\theta) \equiv P_\theta[\underline{Y} = \underline{y}] = P_\theta[Y_1 = y_1, \ldots, Y_n = y_n]$$

Sob independência:

$$L(\theta) \equiv \prod_{i=1}^{n} P_\theta[Y_i = y_i]$$

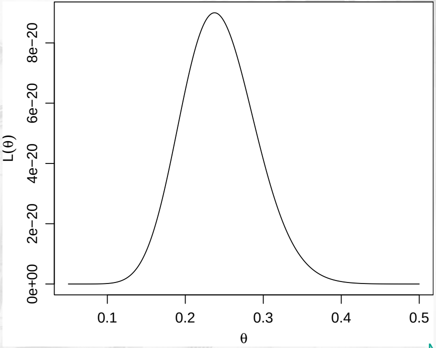
Ou, na escala logarítmica:

$$l(\theta) = \log\{L(\theta)\} \equiv \sum_{i=1}^{n} \log\{P_\theta[Y_i = y_i]\}$$

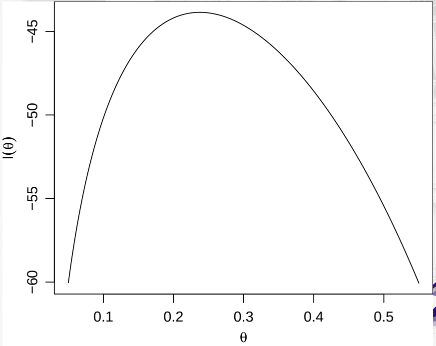
Exemplo 1: estimar proporção de um atributo em uma população

$$Y \sim \mathrm{Bin}(n, \theta) \quad \text{Dados: } (n = 80, y = 19), \text{ amostra aleatória}$$

$$L(\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$$
$$\propto \theta^y(1-\theta)^{n-y} = \theta^{19}(1-\theta)^{80-19}$$



$$l(\theta) \propto y\log\theta + (n-y)\log(1-\theta)$$
$$\propto 19\log\theta + (80-19)\log(1-\theta)$$



Exemplo 1: estimar proporção de um atributo em uma população

Códigos para função de verossimilhança e log-verossimilhança:

```
## função de verossimilhança
L.theta <- function(theta, n, y) theta^y * (1-theta)^(n-y)
##
## função de log-verossimilhança
l.theta <- function(theta, n, y) y * log(theta) + (n-y) * log(1-theta)
```

Código alternativo:

```
L.theta <- function(theta, n, y) dbinom(y, size=n, prob=theta)
##
l.theta <- function(theta, n, y) dbinom(y, size=n, prob=theta, log=TRUE)
```

DEST Departamento de Estatística

Encontrando a estimativa (1)

Analiticamente:

$$\frac{\mathrm{d}}{\mathrm{d}\theta} l(\theta) = 0$$

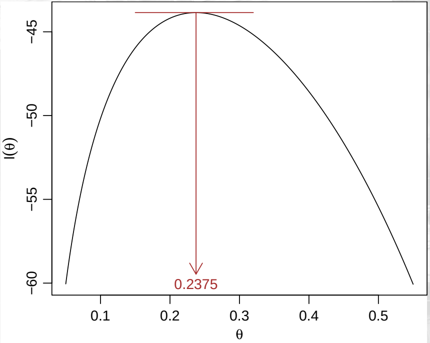
$$\frac{\mathrm{d}}{\mathrm{d}\theta} y \log \hat{\theta} + (n - y) \log(1 - \hat{\theta}) = 0$$

$$\frac{y}{\hat{\theta}} - \frac{n - y}{1 - \hat{\theta}} = 0$$

$$\hat{\theta} = \frac{y}{n}$$

no exemplo:

$$\hat{\theta} = \frac{19}{80} = 0.2375$$



Solução analítica: modelos simples, com poucos parâmetros.
Ilustra conceitos, pouco usada na prática.

Encontrando a estimativa (2)

Numericamente (caso solução analítica não seja possível):

```
(l.max <- optimize(l.theta, n=80, y=19, interval=c(0,1), maximum=TRUE))
## $maximum
## [1] 0.237502
##
## $objective
## [1] -43.85448
```

Solução numérica: aplicável de forma mais geral.
Exige uso (cuidadoso) de algorítmos numéricos.

DEST Departamento de Estatística

Erro padrão

$$s.e.(\hat{\theta}) = \sqrt{-\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} l(\hat{\theta})\right)^{-1}} = \sqrt{-\mathrm{H}(\hat{\theta})^{-1}}$$

Solução analítica para o exemplo:

$$H(\hat{\theta}) = \frac{\mathrm{d}^2}{\mathrm{d}\theta^2} l(\hat{\theta}) = -\left(\frac{y}{\hat{\theta}^2} + \frac{n-y}{(1-\hat{\theta})^2}\right)$$

```
h.theta <- function(theta, n, y) -((y/(theta^2))+((n-y)/((1-theta)^2)))
(H.theta <- h.theta(19/80, y=19, n=80))
## [1] -441.7601
(se.theta <- sqrt(-1/H.theta))
## [1] 0.04757806
## aproximação numérica:
(Hn.theta <- drop(numDeriv::hessian(l.theta, x=l.max$max, n=80, y=19)))
## [1] -441.7551
(sen.theta <- sqrt(-1/Hn.theta))
## [1] 0.04757833
```

Representações alternativas da função de verossimilhança

Verossimilhança Relativa

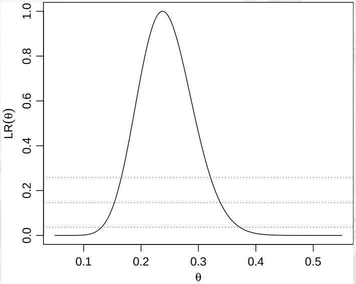
$$\mathrm{LR}(\theta) = \mathrm{L}(\theta)/\mathrm{L}(\hat{\theta})$$

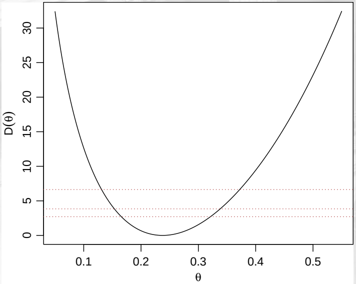
Deviance

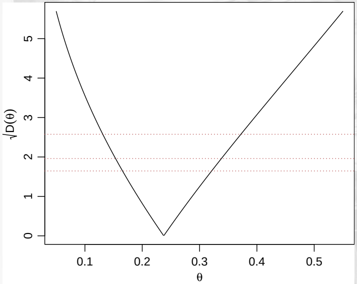
$$\mathrm{D}(\theta) = -2\log\{\mathrm{LR}(\theta)\}$$

Raiz da Deviance

$$\sqrt{\mathrm{D}(\theta)}$$







Exemplo 1 (cont)

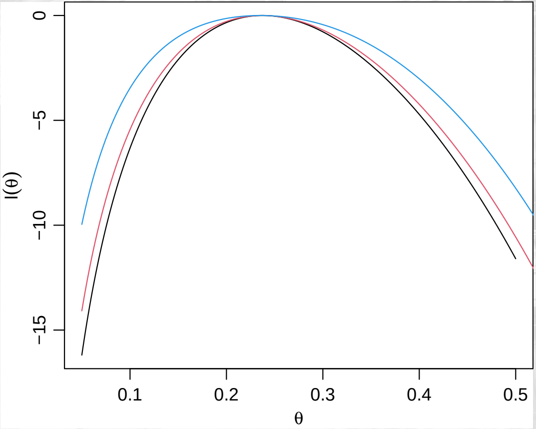
E se o dado fosse um pouco diferente:
Em 80 indivíduos não se sabe exatamente quantos contém o atributo, mas sabe-se que são entre 17 e 21 indivíduos.

Mudaria alguma coisa?

$$L(\theta) \equiv P[17 \leq Y \leq 21] = F(21) - F(16)$$

E se fossem entre 14 e 24 indivíduos?

$$L(\theta) \equiv P[14 \leq Y \leq 24] = F(24) - F(13)$$



Estimativas e erros padrão

$$s.e.(\hat{\theta}) = \sqrt{-\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} l(\hat{\theta})\right)^{-1}} = \sqrt{-\mathrm{H}(\hat{\theta})^{-1}}$$

$y = 19$

```
l.max
## $maximum
## [1] 0.237502
##
## $objective
## [1] -43.85448
se.theta
## [1] 0.04757806
```

$17 \leq y \leq 21$

```
li.max
## $maximum
## [1] 0.2370442
##
## $objective
## [1] -0.7165687
sei.theta
## [1] 0.05102463
```

$14 \leq y \leq 24$

```
li.max2
## $maximum
## [1] 0.2351699
##
## $objective
## [1] -0.1586772
sei2.theta
## [1] 0.06893976
```

- As estimativas podem diferir.
- As log-verossimilhanças (`objective`) não são comparáveis, pois o dado é diferente.
- O erro padrão cresce à medida que o dado se torna mais impreciso.

Códigos

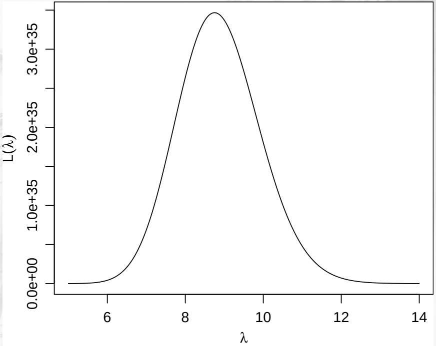
```r
## função de verossimilhança para observação intervalar
li.theta <- function(theta, n, y1, y2) log(pbinom(y2, size=n, prob=theta) -
                                          pbinom(y1-1, size=n, prob=theta))
## encontrando a estimativa: maximizando a função
(li.max <- optimize(li.theta, n=80, y1=17, y2=21, interval=c(0,1), maximum=TRUE))
## $maximum
## [1] 0.2370442
##
## $objective
## [1] -0.7165687
## encontrando o erro padrão
(sei.theta <- sqrt(-1/drop(numDeriv::hessian(li.theta, x=li.max$max, n=80, y1=17, y2=21))))
## [1] 0.05102463
```

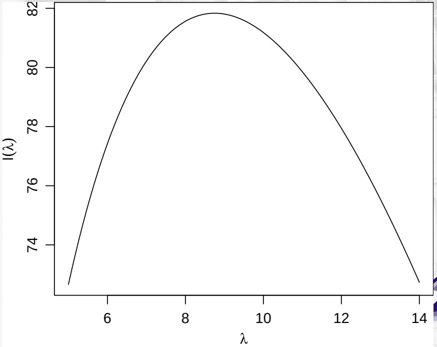
Exemplo 2: estimar parâmetro da distribuição Poisson

Suponha que o número diário de vendas de um produto tem distribuição de Poisson e que em um período obteve-se os valores:

$$12, \quad 7, \quad 8, \quad 5, \quad 11, \quad 9, \quad 10, \quad 8.$$

$$L(\lambda) \propto \exp\{-n\lambda\} \cdot \lambda^{n\overline{y}} \qquad l(\lambda) \propto -n(\lambda - \overline{y} \log \lambda)$$





Exemplo 2: (log) verossimilhança

$$Y \sim P(\lambda)$$

Dados: $(y_1, \ldots, y_n)$, amostra aleatória

$$L(\lambda) = \prod_{i=1}^{n} \frac{\exp\{-\lambda\}\lambda^{y_i}}{y_i!} = \frac{\exp\{-n\lambda\}\lambda^{\sum_{i=1}^{n} y_i}}{\prod_{i=1}^{n} y_i!}$$

$$\propto \exp\{-n\lambda\}\lambda^{\sum_{i=1}^{n} y_i} = \exp\{-n\lambda\}\lambda^{n\bar{y}}$$

$$l(\lambda) = \sum_{i=1}^{n} \{\lambda \; + y_i \cdot \lambda - \log(y_i!)\}$$

$$\propto -n\lambda + \left(\sum_{i=1}^{n} y_i\right) \cdot \log(\lambda) = -n\left(\lambda - \bar{y} \cdot \log(\lambda)\right)$$

Exemplo 2: verossimilhança para λ no modelo Poisson

Códigos para função de verossimilhança e log-verossimilhança:

```
## função de verossimilhança
L.lambda <- function(lambda, n, soma) exp(-n*lambda) * lambda^soma
##
## função de log-verossimilhança
l.lambda <- function(lambda, n, soma) -n*lambda +  soma * log(lambda)
```

Código alternativo:

```
## função de verossimilhança
L.lambda <- function(lambda, dados) prod(dpois(y, size=n, prob=lambda))
##
## função de log-verossimilhança
l.lambda <- Vectorize(function(lambda, dados) sum(dpois(y, lambda=lambda, log=TRUE)), "lambda")
```

DEST Departamento de Estatística

Continuando …

Fazer por analogia com problama anterior.

- Funções escore e hessiano.
- Soluções analítica e numérica.
- Erros padrão e intervalos.
- Função de verossimilhança relativa, *deviance* e raiz da *deviance*.
- Observações *imprecisas*.

Expressão da Verossimilhança (II)

V.A. contínua: medição à certa precisão $(y_{iI} \leq y_i \leq y_{iS})$

- Forma mais geral (densidade multivariada)

$$L(\theta) = P_\theta[y_{1I} \leq Y_1 \leq y_{1S}, y_{2I} \leq Y_2 \leq y_{2S}, \ldots, y_{nI} \leq Y_n \leq y_{nS}]$$

- Sob independência

$$\begin{aligned} L(\theta) &= P_\theta[y_{1I} \leq Y_1 \leq y_{1S}] \cdot P_\theta[y_{2I} \leq Y_2 \leq y_{2S}] \ldots P_\theta[y_{nI} \leq Y_n \leq y_{nS}] \\ &= \prod_{i=1}^{n} P_\theta[y_{iI} \leq Y_i \leq y_{iS}] \end{aligned}$$

- Se grau de precisão comum, $(y_i - \delta/2 \leq Y_i \leq y_i + \delta/2)$;

$$L(\theta) = \prod_{i=1}^{n} P_\theta[y_i - \delta/2 \leq Y_i \leq y_i + \delta/2] = \prod_{i=1}^{n} \int_{y_i-\delta/2}^{y_i+\delta/2} f(y_i, \underline{\theta}) \mathrm{d}y_i.$$

Expressão da Verossimilhança (II) (cont)

- alto grau de precisão (δ é pequeno em relação a variabilidade dos dados)

$$L(\theta) \approx \left(\prod_{i=1}^{n} f(y_i, \underline{\theta}) \right) \delta^n,$$

- e se δ não depende dos valores dos parâmetros

$$L(\theta) \approx \prod_{i=1}^{n} f(y_i, \underline{\theta}) \text{ e } l(\theta) \approx \sum_{i=1}^{n} \log \{ f(y_i, \underline{\theta}) \}$$

- observações não independentes - densidade multivariada:

$$L(\theta) = f(\underline{y}, \underline{\theta}) \text{ e } l(\theta) \approx \log \{ f(\underline{y}, \underline{\theta}) \}$$

Exemplo 3: Estimando uma média

Deseja-se estimar a vazão média de um rio. Foram tomadas três medidas (em dm^3/s):

$$y_1 = 20, \quad y_2 = 28, \quad y_3 = 39$$

Uma estimativa da média é obtida por:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{20 + 28 + 39}{3} = 29.0$$

Podemos expressar incerteza reportando um intervalo de confiança:

$$\bar{y} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

O intervalo (95% de confiança) então é (vamos supor $\sigma = 5$):

$$29.0 \pm 1.96 \frac{5}{\sqrt{3}} \longrightarrow 29.0 \pm 5.7$$

(23.3 ; 34.7)

Estimando a média

Deseja-se estimar a vazão média de um rio. Foram tomadas três medidas (em dm^3/s):

$$y_1 > 20, \quad 22 < y_2 < 30, \quad y_3 < 35$$

Uma estimativa da média é obtida por:

$$???$$

Podemos expressar incerteza reportando um intervalo de confiança:

$$???$$

- A solução não é óbvia,
- mas a pergunta ainda faz sentido!

DEST Departamento de Estatística

Recomeçando

Deseja-se estimar a vazão média de um rio. Foram tomadas três medidas (em dm^3/s):

$$y_1 = 20, \quad y_2 = 28, \quad y_3 = 39$$

Uma estimativa da média é obtida por ...

Qual o valor $\hat{\mu}$ para o qual se tem a maior chance de observar

$$P[(Y_1 = 20) \cap (Y_2 = 28) \cap (Y_3 = 39)|\mu]$$

Equivalentemente, qual o valor $\hat{\mu}$ mais compatível com os dados

$$P[(Y_1 = 20) \cap (Y_2 = 28) \cap (Y_3 = 39)|\mu]$$

Assumindo $Y \sim \mathrm{N}(\mu, \sigma^2 = 5^2)$

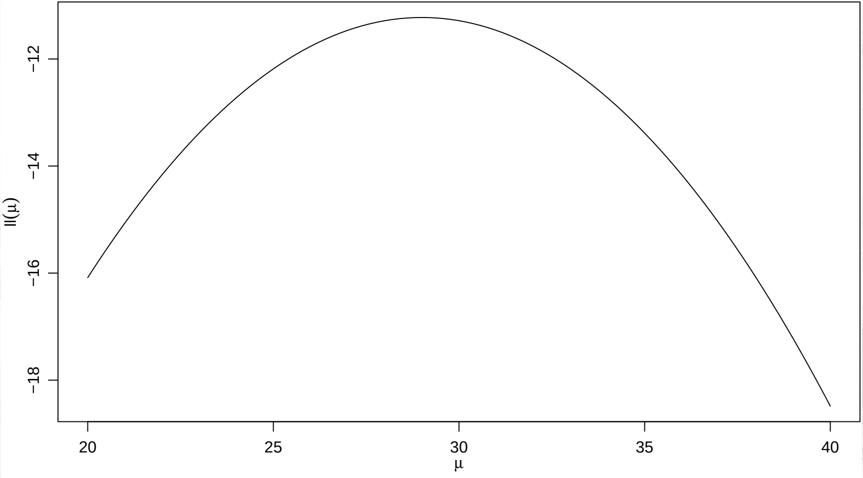
$$L(\mu) = P[(Y_1 = 20) \cap (Y_2 = 28) \cap (Y_3 = 39)|\mu]$$

Função de verossimilhança!!!

DEST Departamento de Estatística

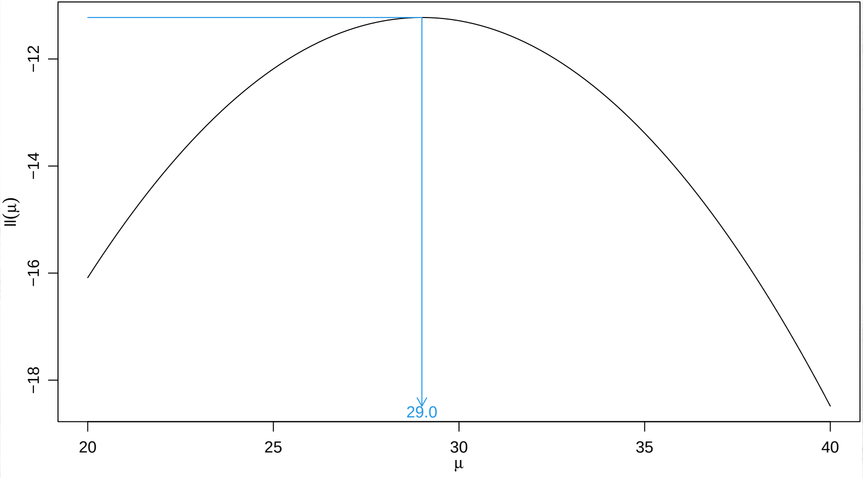
Revisitando o cálculo da média

$$y_1 = 20, \quad y_2 = 28, \quad y_3 = 39$$



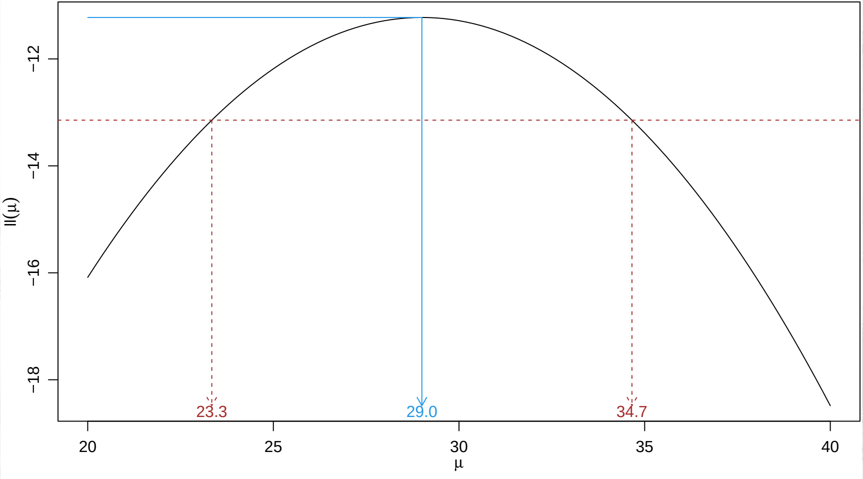
Revisitando o cálculo da média

$$y_1 = 20, \quad y_2 = 28, \quad y_3 = 39$$



Revisitando o cálculo da média

$$y_1 = 20, \quad y_2 = 28, \quad y_3 = 39$$



Retomando

Deseja-se estimar a vazão média de um rio. Foram tomadas três medidas (em dm^3/s):

$$y_1 > 20, \quad 22 < y_2 < 30, \quad y_3 < 35$$

Uma estimativa da média é obtida por:

???

E podemos expressar incerteza reportando um intervalo de confiança

???

DEST Departamento de Estatística

Retomando

Deseja-se estimar a vazão média de um rio. Foram tomadas três medidas (em dm^3/s):

$$y_1 > 20, \quad 22 < y_2 < 30, \quad y_3 < 35$$

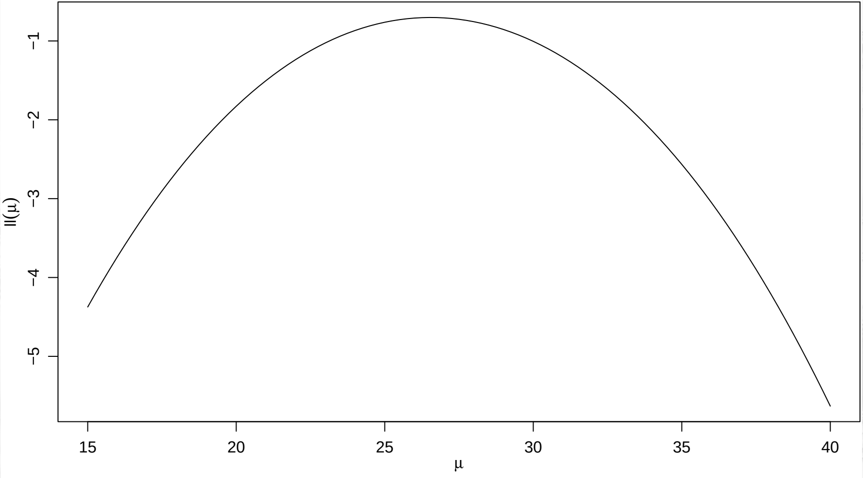
Uma estimativa da média é obtida por ...
encontrar $\hat{\mu}$ que maximiza

$$L(\mu) = P[(Y_1 < 20) \cap (22 < Y_2 < 30) \cap (Y_3 < 35)|\mu]$$

E podemos expressar incerteza reportando um intervalo de confiança ...
...dado por um corte que define uma faixa de valores *aceitáveis* na função de verossimilhança

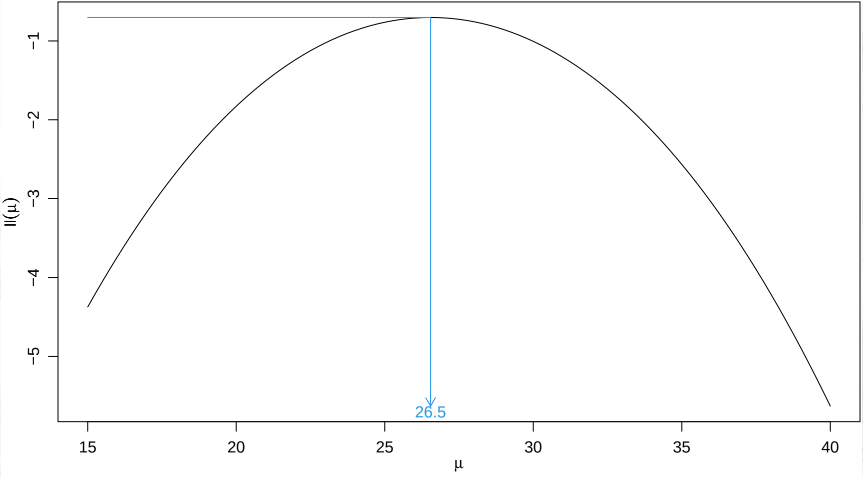
Resolvendo o cálculo da média

$$y_1 > 20, \quad 22 < y_2 < 30, \quad y_3 < 35$$



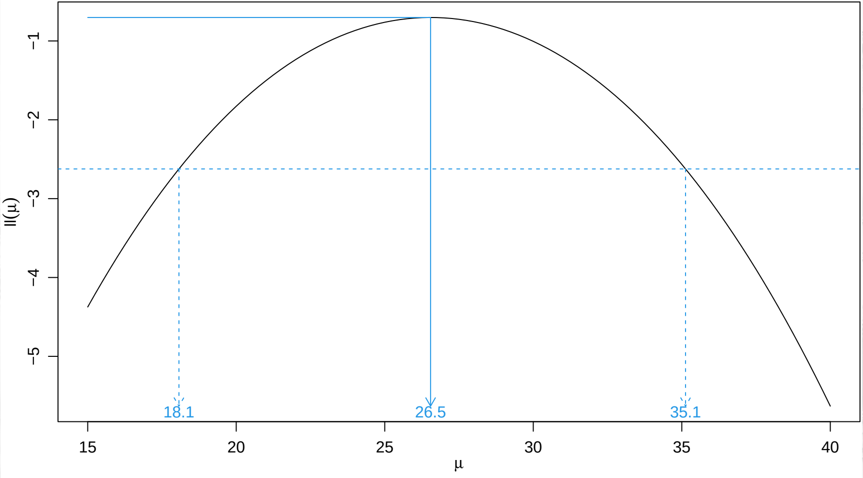
Resolvendo o cálculo da média

$$y_1 > 20, \quad 22 < y_2 < 30, \quad y_3 < 35$$



Resolvendo o cálculo da média

$$y_1 > 20, \quad 22 < y_2 < 30, \quad y_3 < 35$$



Dados intervalares

Dados intervalares ou censurados

- Terminologia
  - $y_1 > 20$: censura à direita
  - $22 < y_2 < 30$: censura intervalar
  - $y_3 < 35$: censura à esquerda
- Generaliza-se para modelos de regressão e outros
- Usuais em análise de sobrevivência

Será que você já teve ou tem dados intervalares?

DEST Departamento de Estatística

Códigos

Média aritmética

```r
y1 <- c(20, 28, 39)
m.y1 <- mean(y1)
```

Função de verossimilhança para estimar média

```r
llint <- Vectorize(
    function(par, y.p, y.i){
        ## y.p é um vetor de dados pontuais
        ## i.i é uma matriz (n x 2) de dados intervalares
        if(missing(y.p)) ll.p <- 0
        else ll.p <-  sum(dnorm(y.p, mean=par, sd=5, log=TRUE))
        if(missing(y.i)) ll.i <- 0
        else ll.i <- sum(apply(y.i, 1, function(x) log(diff(pnorm(x, mean=par, sd=5)))))
        return(ll.i+ll.p)
    },
    "par")
```

Códigos

Dados pontuais:

```
## a) 20, 28, 39
#y1 <- c(20, 28, 39)
#m.y1 <- mean(y1)
##curve(llint(x, y.p = y1), from=20, to=40)
mu.vals <- seq(20,40,l=101)
ll.y1 <- llint(mu.vals, y.p = y1)
fit.y1 <- optimize(llint, int=c(20,40), y.p=y1, maximum=TRUE)
llcut <- function(par, llfun, fit, level=0.95, ...){
    X2 <- qchisq(level, df=1)
    with(fit, -2*(llfun(par, ...) - objective) - X2)
}
require(rootSolve)
IC1 <- uniroot.all(llcut, interval=c(20,40), llfun=llint,fit=fit.y1, y.p=y1)
IC1exato <- mean(y1) + c(-1,1) * 1.96 * 5/sqrt(length(y1))
corte <- with(fit.y1, objective - qchisq(0.95, df=1)/2)
```

Códigos

Dados Intervalares

```
## Dados:  >20,  [22-30] , <35
##(y2  <- cbind(c(22, 25, -Inf), c(Inf, 32, 35)))
(y2  <- cbind(c(20, 22, -Inf), c(Inf, 30, 35)))
##      [,1] [,2]
## [1,]   20  Inf
## [2,]   22   30
## [3,] -Inf   35
mu.vals <- seq(15,40,l=101)
ll.y2 <- llint(mu.vals, y.i = y2)

fit.y2 <- optimize(llint, int=c(20,40), y.i=y2, maximum=TRUE)
IC2 <- uniroot.all(llcut, interval=c(10,40), llfun=llint,fit=fit.y2, y.i=y2)
corte <- with(fit.y2, objective - qchisq(0.95, df=1)/2)
```

DEST Departamento de Estatística

Um outro exemplo

Dados (amostra):

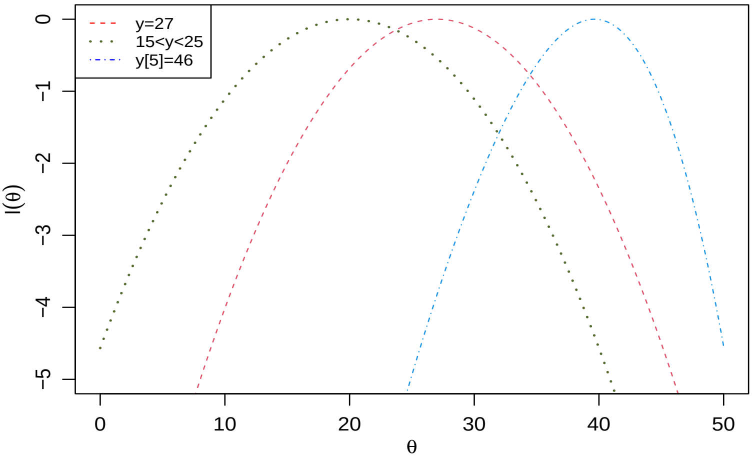
- $y = 27$ (uma observação é 27)
- $15 < y < 25$ (uma observação está entre 15 e 25)
- máximo $y_{[5]} = 46$ (dentre cinco observações 46 é a maior)

E ainda queremos responder às mesmas questões

- Qual a média?
- Qual a variabilidade?
- A média estima o valor central de uma população?
- Com qual incerteza?
- Há evidências para afirmar que a média (da população) difere de 38?

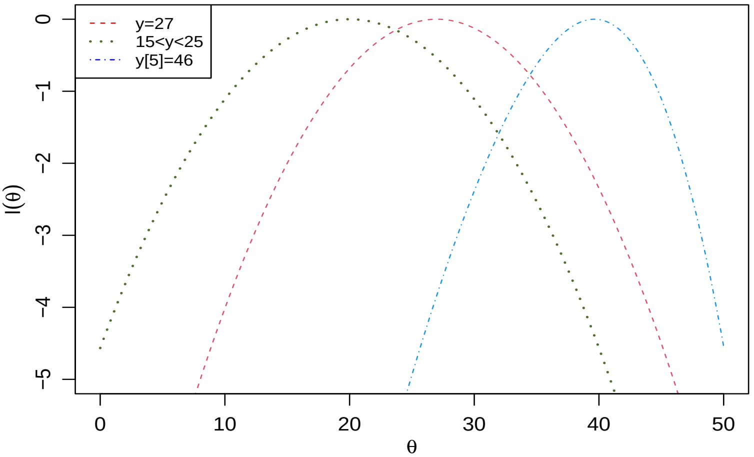
A verossimilhança (e informação) de cada dado

$$15 < y < 25 \quad ; \quad y = 27 \quad ; \quad y_{[5]} = 46$$



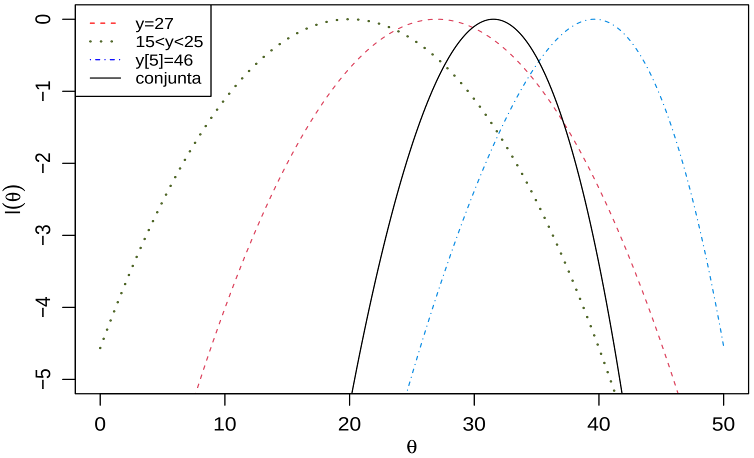
A verossimilhança (e informação) de cada dado

$$15 < y < 25 \quad ; \quad y = 27 \quad ; \quad y_{[5]} = 46$$

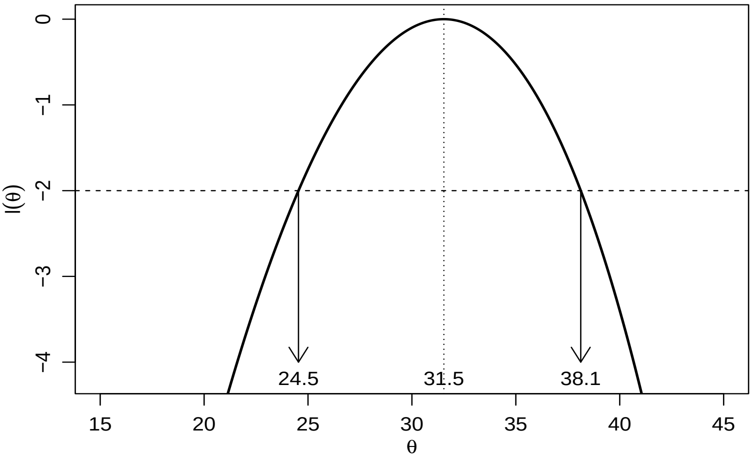


Combinando as informações de cada dado

$$15 < y < 25 \quad ; \quad y = 27 \quad ; \quad y_{[5]} = 46$$



Extraindo informações de interesse



Resumindo

- A função de verossimilhança contém informação do dado sobre parâmetro(s) do modelo.
- Permite obter estimativa(s) com respectiva incerteza e avaliar hipóteses sobre parâmetro(s).
- Para modelos mais gerais, a função pode ser complexa e algorítmos numéricos são necessários.
- Variações, adaptações e formas alterantivas (pseudo-verossimilhanças) podem ser usadas, mas mantêm a mesma intuição de resumir a informação da amostra.
- A função pode fornecer o estimador para métodos frequentistas e também é central na abordagem Bayesiana.

DEST Departamento de Estatística